Registering Oblique Image Pairs Without Camera Models

Robert M. Taylor Jr.
Johns Hopkins University
Applied Physics Laboratory

24 May 1999

Abstract

This paper presents automatic electro-optical image registration of images taken from oblique views (large baseline separation). The method uses feature based point correspondence to solve for the projective parameters in an iterative approach. The feature points are solved for by analyzing region contours on segmented images. A simple geometric invariance method is used to match points and remove outliers. A projective transform is applied to previously warped images until the tie point correspondence satisfies a suitable error measure. The series of projective transformation parameters are then combined to solve for the overall global transformation.

Work completed under JHU/APL Internal Research and Development

Form SF298 Citation Data

Report Date *("DD MON YYYY")* 24051999	**Report Type** N/A	**Dates Covered (from... to)** *("DD MON YYYY")*

Title and Subtitle Registering Oblique Image Pairs Without Camera Models Applications	**Contract or Grant Number**
	Program Element Number
Authors Taylor, Robert M.	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Johns Hopkins University Applied Physics Laboratory	**Performing Organization Number(s)**
Sponsoring/Monitoring Agency Name(s) and Address(es)	**Monitoring Agency Acronym**
	Monitoring Agency Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited	
Supplementary Notes	
Abstract	
Subject Terms	
Document Classification unclassified	**Classification of SF298** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** unlimited
Number of Pages 17	

1 Introduction

Research in image alignment has yet to uncover a robust general solution that will solve the problem of aligning arbitrary perspective aerial images in the absence of camera models. Camera models are often included in image headers and contain parameters or equations that relate the ground surface to the image plane. In addition to the use of sensor data as used in [2], other forms of support data such as controlled imagery databases, elevation models, and site models [8] are often used to aid the registration process. Zheng and Chellappa [9] use a parameter estimation scheme for working the oblique registration problem. The techniques proposed here are unique in that a feature based scheme is used for finding 2-D tie points on a 3-D surface viewed from arbitrary perspectives without the use of parameter estimation or camera models. Previous work using feature point methods rely on images being related by rigid or affine transformations. The eight projective parameters are computed merely from the point correspondence established entirely by simple low-level image processing. For a comprehensive look at some of the general methods in image registration see Brown [1]. Applications of this process include coordinate mensuration for targeting, mosaicing, depth recovery, camera motion estimation, and data fusion with other sensing modalities.

1.1 Scope of Problem

This paper focuses on the problem of performing automatic alignment on multi-perspective electro-optical images with no support information. The problem domain is limited to aligning aerial photographs. With this constraint, the ground surface in the field of view is roughly a planar patch. This will allow for a global transformation to remove perspective distortions and alleviate the need for locally varying transformation models. Once the global transformation is performed, locally varying translational models can be used to remove effects of hilly terrain. Local transformations and adjustments for subpixel accuracy are not discussed here. In general, for large oblique angles, it is difficult to obtain a standard for subpixel accuracy because the warping transformation will always introduce new information through interpolation or remove information through decimation. The assumptions made about the nature of the image pairs to be registered are summarized as follows:

- Contents of images lie on a plane (effects of varying elevation can be ignored)
- Majority of image plane contains common scene content between images
- Scene contains regions that can be segmented

Note that there is no restriction on lighting conditions between image pairs or on the date of collection. The images could be taken years apart as long as there are still common components to the scene. Figure 1 shows a typical collection picture for acquiring the images to be registered. Again, no a priori knowledge of the position or orientation of the platforms is required.

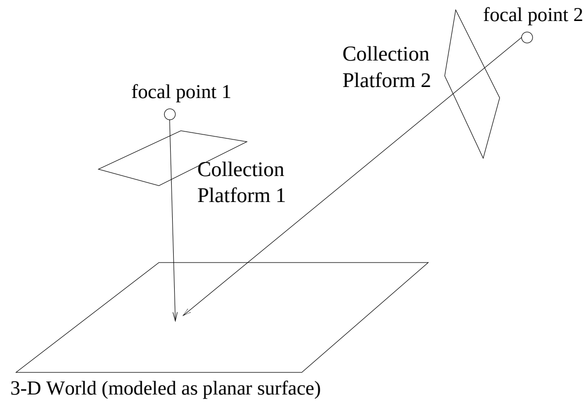

Figure 1: Sketch of a typical image collection scenario where the positions and orientations of the sensing platforms is unknown. There is no requirement for the sensors to be in close proximity either in time or space. The only input is the image pair itself.

1.2 General Approach

Figure 2 shows a flowchart of the steps involved in registering two oblique image views. Since the images can vary in intensity and information content, we must rely on shape features to perform the correspondence. The first step in the process is to segment both images. Next, edge detection on the segmented regions is used to pull out boundaries. The third step is to process these boundaries to form an array of 1-D contour descriptors. These arrays are processed in the fourth step to identify candidate feature points. Some candidate points are discarded as part of this step. In the fifth step, the candidate feature points are matched. Matching points are used to determine the projective transformation between the images, and one image is warped to the other. This warping is not perfect, and the process is repeated until a suitable error criterion is satisfied.

Section 2 covers the conversion from images to objects to boundaries to boundary descriptors. Section 3 will discusses method used for feature point extraction. Section 4 shows how the feature points are matched for correspondence and how false matching pairs are eliminated. Section 5 presents the projective warping method used in the final step of each iteration and the computation of the error. Section 6 discusses results on an actual image pair, and Section 7 presents a summary.

2 Segmentation and Boundary Extraction

Figure 3 shows the original image pair that serves as the input to the algorithm. To define boundaries of objects we need a method that will perform texture segmentation to discriminate between the border of an object and high frequency patterns on the object itself. Here the use of the word “object” applies to any portion of the scene that is statistically similar. Figure 4 shows the image pair after application of a texture segmentation algorithm. Any one of a number of segmentation routines may be used; a variation of the watershed algorithm was used here. Figure 5 shows the image pair after application of an edge detector applied to the segmented image. A simple sobel edge detector was applied here.

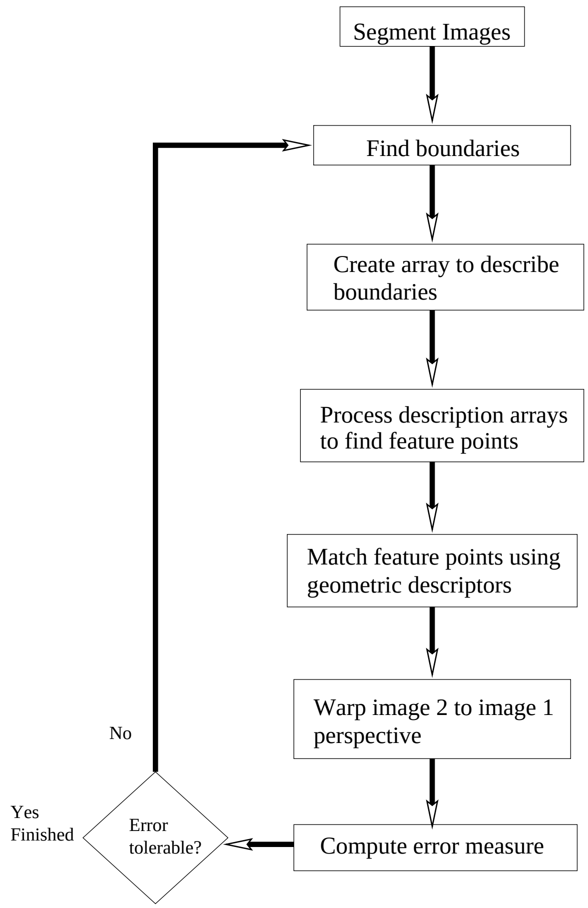

Figure 2: Flowchart for oblique image registration.

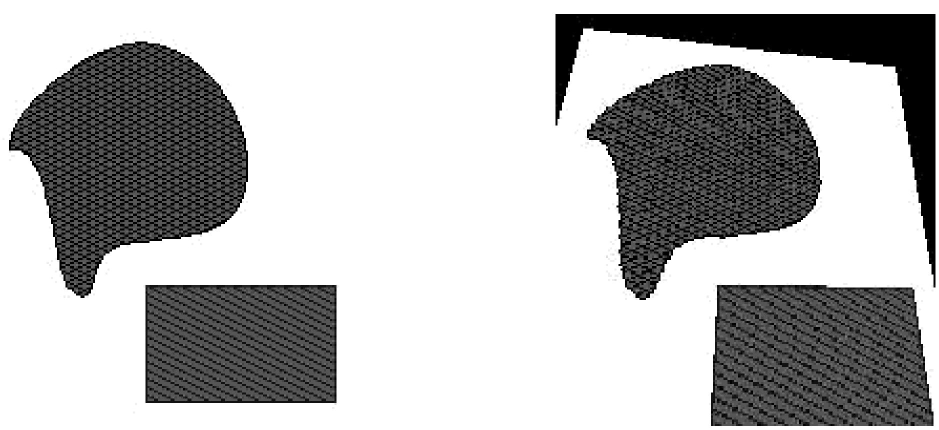

Figure 3: Synthetic image pair used to illustrate concepts of registration algorithm. The two images can be related by a global projective transformation.

Figure 4: Texture segmented images

Figure 5: Edge images created from application of edge detector to the segmented images. In this way, only the object boundaries are chosen, not high frequency texture information.

3 Feature Extraction

The feature extraction phase is the most critical stage in the algorithm because a minimum of four feature points in common between the images is required. Without four or more matching points, the projective parameters cannot be solved for. We have chosen the feature points to be points of maximum curvature (or minimum radius of curvature). For scenes containing corners this will become a corner selector.

3.1 Polar Contour Extraction

The features chosen to be extracted are points that lie on or near a point of maximum curvature in the edge image. To find these points of maximum curvature, we characterize the contours in the edge image with 1-D polar plots of distance and angle from a particular pivot index in the image called a seed. Once a seed is established, all points in the vicinity of the seed can be parameterized by either a Cartesian vector or polar parameters. Letting the angle $\theta \in [0, 2\pi]$ plots of ρ versus θ for each seed can be generated. Of course the difficult part is knowing which contour to extract near a given seed. The convention is to select the closest contour that exhibits the greatest amount of continuity over the angle domain. In other words, trace out the most complete boundary that encircles a given point. A pictorial description of this is given in Figures 6 and 7. The figures show how the polar plot varies as a function of seed location. However, even though the curves vary, distinct peaks (both local and global) can be found that correspond to points of maximum curvature. In the case of a rectangle the four local peaks map to the corners. For the amorphous shape local peaks in the 1-D plots represent the sharpest bends in the contour.

As can be seen, the location of the seeds is important for determining the polar plots. Ideally, the seed to lie at the centroid of a closed contour. In practice this is difficult since closed contours don't always exist. Usually, contours are not closed due to object occlusion, poor texture segmentation, or a scene content that does not contain confined separated regions. What is desired instead is to look at placing seeds in locations where no there are no boundaries within a certain distance. If seeds are placed too close to boundaries, the polar plot will not correctly reveal all feature points. One way to place seeds in regions where there are no boundaries is to look at the quadtree decomposition of the edge image. The quadtree decomposition systematically decomposes an image into square blocks sized by a power of two where the criteria for decomposing in this case is set to true if an edge pixel is found in a given decomposition block. The largest blocks in the decomposition serve as markers for where to place seeds because they represent regions where there are no edge pixels within that particular block area. Figure 8 shows the results of applying the quadtree decomposition to the synthetic images.

3.2 Processing the Polar Contours

Having established the polar plot, we can determine points of minimum radius of curvature by looking for local peaks in the polar plot. Note from the polar plots in Figures 6 and 7 show that the arrays are continuous and have few high frequency components. That is because we applied the polar contour extraction method to an ideal closed contour. In general, there will be gaps or holes in the boundaries. Furthermore, there will be high

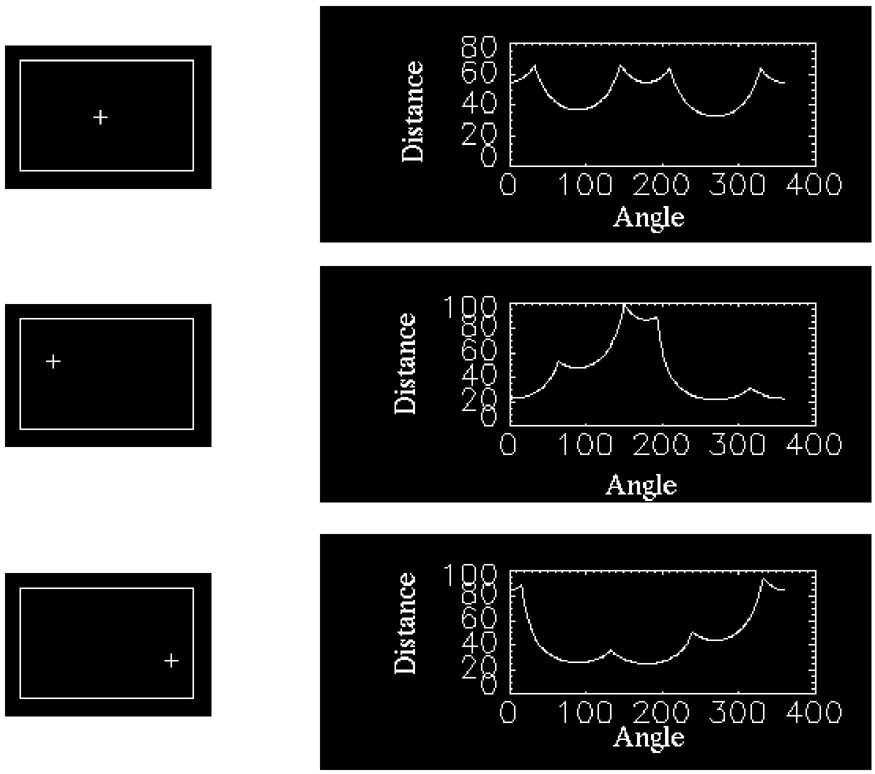

Figure 6: Rectangle shape: The images on the left side show the convex boundaries and seed location which give rise to the corresponding polar plots displayed on the right side. The local peaks or points of discontinuity represent candidates for feature points (corner points in this case).

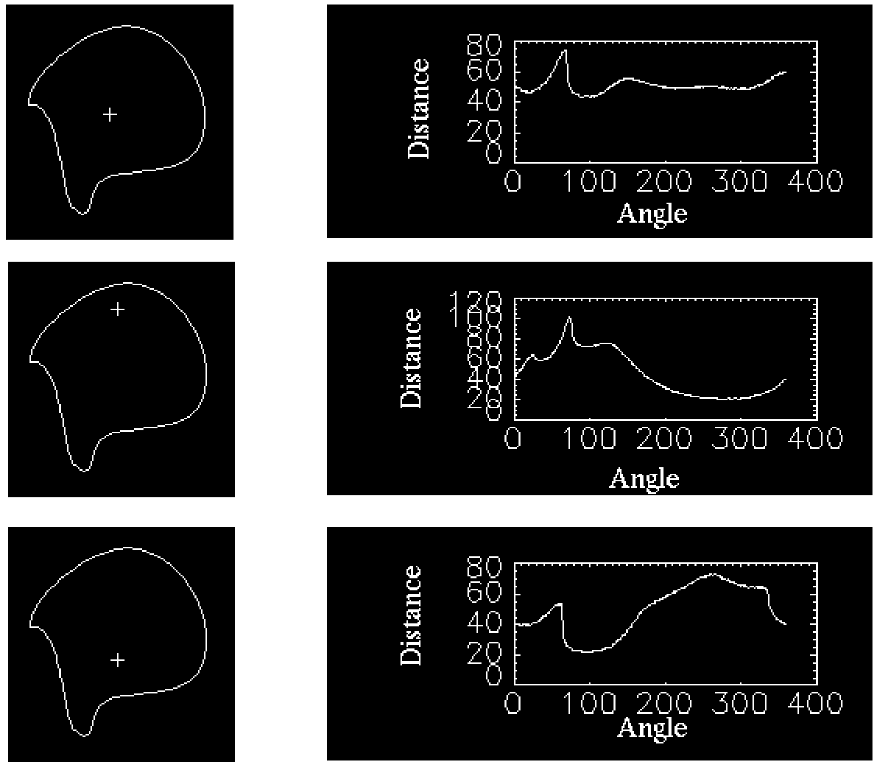

Figure 7: Arbitrary shape:The images on the left side show the convex boundaries and seed location which give rise to the corresponding polar plots displayed on the right side. Here, one very well resolved spike gives the point of maximum curvature. Other local peaks show locations of smaller curvature in the boundary.

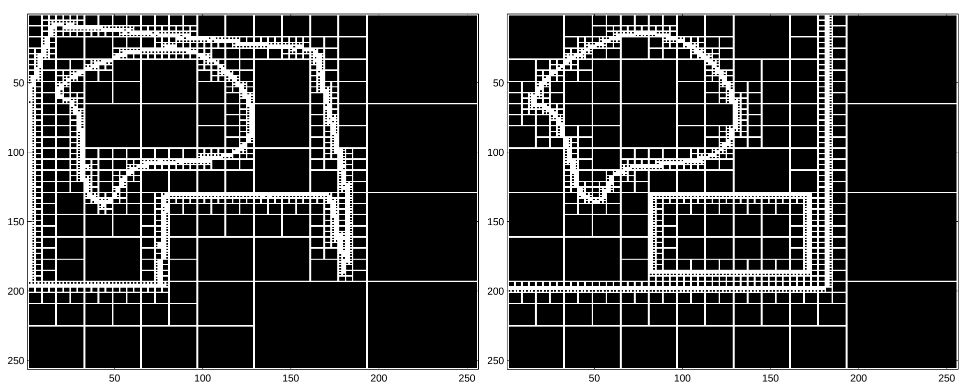

Figure 8: Quadtree decomposition block image for the synthetic images. Note that the largest blocks lie in regions where there are no boundary pixels. These regions will be used for planting seeds.

frequency spikes in the polar plot coming from points that were grabbed outside of the closest contour to the seed. These problems can be alleviated by smoothing the 1-D polar description arrays with median filters or low pass filters. The next step is to find points of zero derivative and negative concavity to locate the positions of all local peaks. Once the peak locations are established in the polar plots, we have ρ and θ for the given peak and the Cartesian location of the seed. From these pieces of information we can solve for the image coordinate that maps to the point of maximum curvature. We must process in like manner all the polar contours that come back for each seed location. If several seeds fall within the same convex boundary, then we can correlate the results of each extracted feature point and throw away points that did not match for all seeds. This is the first layer of feature point removal. The second layer comes during the matching stage.

4 Matching Feature Points

Once the feature points have been selected, the next step is to match common points and reject non-matching points. The only basis we had for selecting the feature points was shape information. We could not rely on texture information in the feature extraction phase because the image pairs in general will have a high degree of intensity variability. We furthermore cannot use texture in determining which points have correspondence even if we look at statistical properties of neighboring pixels. Moment invariance can be removed as a matching criterion because we are dealing with projective distortion between images. Correlation techniques are infeasible due to the unknown spatial distortion (i.e. we do not know a priori rotation, scale, skew, or any projective parameters). This leaves only one option to exploit–shape information.

As illustrated in the next section, for 2-D rigid transformations in the image plane, angle magnitudes and ratios of line lengths are preserved. While not precisely true for projective transformation, changes in angle magnitudes and length ratios are less sensitive to differences in viewing angles than some other similarity measures. Thus these geometrical constraints will be used to determine point correspondence between images.

4.1 Geometric Invariance

Given a set of feature points for each image, we must be able to extract features from the two sets that will determine the projective transformation. If we first consider the special case of a 2-D rigid body transformation in the image plane, the matching points will be related by only three parameters: rotation, scale, and translation. As Figure 9 shows, the angles between points and the ratios of distances between points will be the same under a rigid body model. Non-matching points will be determined as those for which no matching angle or distance ratio was found.

We can relax the constraint that the feature sets be related by a rigid transformation to instead a projective transformation. This will entail creating thresholds for an allowable difference in angle and distance ratios. The geometric constraint problem could now be described by relaxed constraint equations such as

$$| \alpha_1 - \alpha_2 |^n < \epsilon, n = 1, 2$$

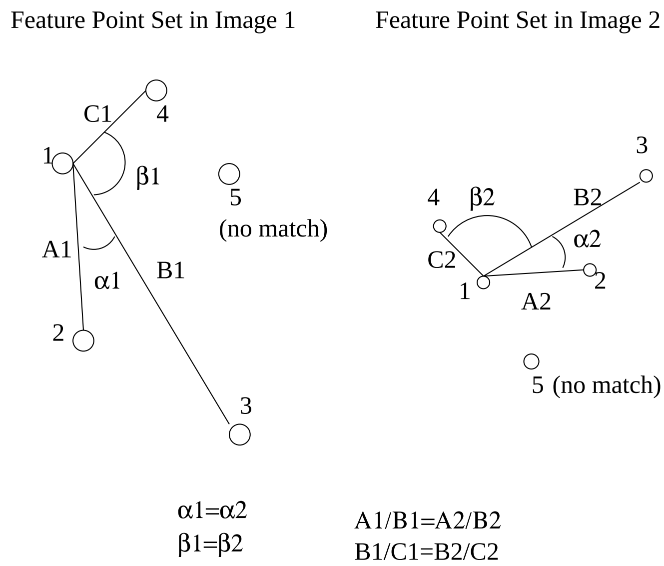

Figure 9: Example feature points determined for each image are shown here with circles. The set of feature points on the right side show the effects of a rotation, scale, and translation being applied toward the points on the left. The angles and line length ratios are preserved however.

$$\mid \frac{A_1}{B_1} - \frac{A_2}{B_2} \mid^n < \epsilon, n = 1, 2 \tag{1}$$

where the A's and B's represent line lengths, the $\alpha's$ and $\beta's$ are the angles between the various lines, and ϵ is a threshold for allowable differences. By comparing sequences of angles and line ratios, we can match individual points between the two feature sets. It is the processing of the sequential angles and ratios that overcomes the angular disparity introduced by the relaxation technique. Lines in the geometrical description are not related to and do not imply that the original images contained lines. The images have been reduced to a list of feature points. All processing from this point is done on number sets, which greatly reduces the computational burden of processing image data and drastically improves speed.

4.2 Finding Matching Points and Removing Outliers

For each point in the feature set we must find all angles and line lengths to every other point in the feature set. A point in feature set 1 is declared to have a match with a point in feature set 2 if the number of sequentially matching angles and line ratios meets a certain threshold. Points that do not meet this criteria are thrown out of the set. When the matching procedure has been applied to every point in both feature sets, the point correspondence has been established. Once the point correspondence has been established, we can proceed to solve for the projective parameters as described in the next section.

5 Projective Warping

It is well known [5] that a point correspondence between two images related by a projective transformation is given by

$$\begin{aligned} x' &= \frac{a_1 x + a_2 y + a_3}{a_7 x + a_8 y + 1} \\ y' &= \frac{a_4 x + a_5 y + a_6}{a_7 x + a_8 y + 1} \end{aligned} \tag{2}$$

where $[x'y']^T$ is a point in image 2 corresponding to a point $[xy]^T$ in image 1, and $a_1 ... a_8$ are the eight projective parameters. We can hold one image fixed and bring the other image into alignment by means of a projective transformation if we can safely assume that we are viewing planar surfaces using a conventional pinhole camera model. The image to be warped into alignment will be called the warping image (or distorted image) and the fixed image the reference. If we collect a new set of feature points on the warped image that correspond to points on the reference image, we can find a new set of projective parameters $b_1 ... b_8$ that map warp image 2 to warp image 1 according to the same equation structure as in 2,

$$\begin{aligned} x'' &= \frac{b_1 x' + b_2 y' + b_3}{b_7 x' + b_8 y' + 1} \\ y'' &= \frac{b_4 x' + b_5 y' + b_6}{b_7 x' + b_8 y' + 1} \end{aligned} \tag{3}$$

where $[x''y'']^T$ is a point in warp image 2 corresponding to a point $[x'y']^T$ in warp image 1, and $b_1 ... b_8$ are the eight new projective parameters relating the newest warped image to the previously warped image. By substituting 2 into 3, we can arrive at a new set of projective coefficients that relate points in warp image 2 to points in the reference image. The new set of eight coefficients $c_1 ... c_8$ can be found through a simple matrix multiplication as follows

$$\begin{bmatrix} c_1 & c_2 & c_3 \\ c_4 & c_5 & c_6 \\ c_7 & c_8 & 1 \end{bmatrix} = \frac{\begin{bmatrix} b_1 & b_2 & b_3 \\ b_4 & b_5 & b_6 \\ b_7 & b_8 & 1 \end{bmatrix} \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & 1 \end{bmatrix}}{\begin{bmatrix} b_7 & b_8 & 1 \end{bmatrix} \begin{bmatrix} a_3 \\ a_6 \\ 1 \end{bmatrix}} \tag{4}$$

In this way, we can keep track of individual projective parameters for each iteration and then solve for the overall projective parameter set that aligns the two original images.

5.1 Warping a Synthetic Image

To experimentally validate the success of the algorithm, a synthetically derived image pair was created by applying a known projective transformation. Figure 10 shows the original and warped synthetic image having undergone the known projective polynomial model. The automatically derived tie points are labeled on the images. From the tie point corre-

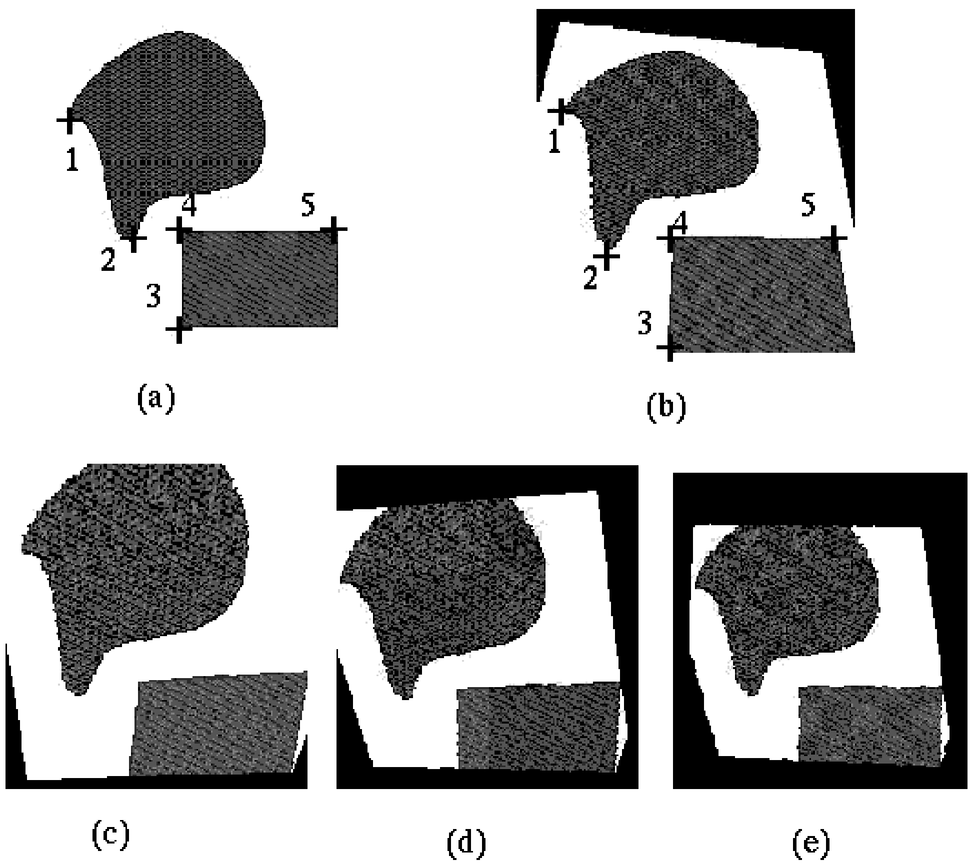

Figure 10: Top row (a,b) shows the original images with the the first pass set of tie points. Bottom row (c,d,e) shows successive stages of the algorithm and the projective warping result after three iterations.

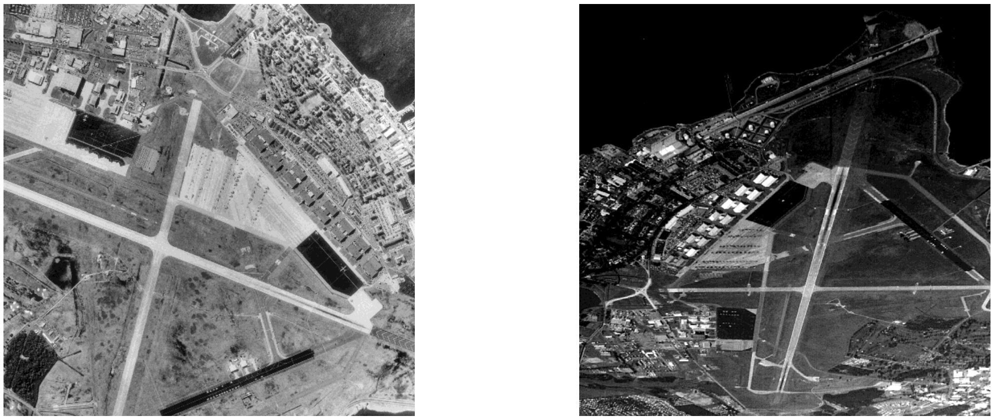

Figure 11: Input image pair is two oblique views of same runway taken months apart. Notice the differences in the information content.

spondence we can now solve for the eight projective parameters that are used to warp the warped image back into alignment with the reference image through a series of iterations. There are several reasons why the projective transformation does not converge after just one iteration. Primarily, the reason for a non-exact transformation lies in inaccuracies in the point correspondence. The point correspondence becomes increasingly better after each iteration of the algorithm. Note that in part (e) of Figure 10 the distorted image is very close in alignment with the reference image.

5.2 Computing a Measure of Fit

The measure for determining the error in the alignment is based on solving for the root mean square (rms) error of all the tie points. To compute the rms error for the precision of the final tie points, use the following formula:

$$rms = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}[(x_{i,r} - x_{i,d})^2 + (y_{i,r} - y_{i,d})^2]} \tag{5}$$

where the subscript "r" represents the reference image,"d" the distorted image, and N the number of point correspondences. Once the iterated algorithm converges to a suitable tie point rms, the procedure stops and the comprehensive projective set of parameters based on equation 4 is computed.

6 Results on an actual image pair

To test the algorithm on an actual set of imagery, we examine two oblique views of a runway taken from different platforms at very different times. In fact, some of the scene content has changed during the time of acquisition. Figure 11 shows the original image pair that serves as the input to the algorithm. Figure 12 shows the image pair after application of a gradient watershed texture segmentation algorithm. Figure 13 shows the image pair after

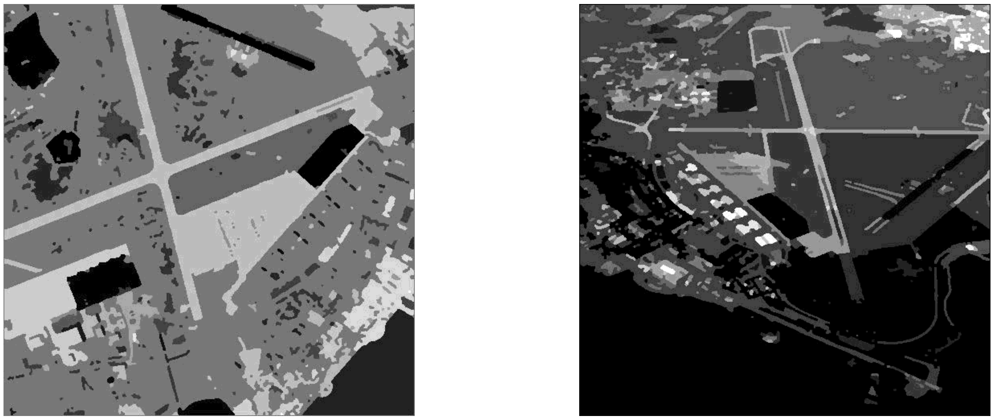

Figure 12: Segmented image pair ready for edge detection. By segmenting the images we remove high frequency components and focus on the real boundary of regions.

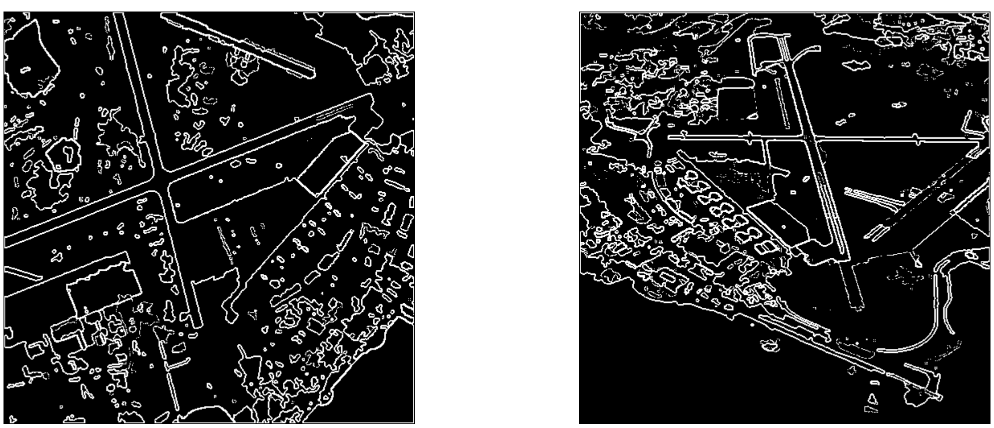

Figure 13: Edge images obtained from application of Canny edge detector to segmented images.

application of a Canny edge detector applied to the segmented image. Figure 14 shows two iterations of the algorithm being applied with the output warped image at the end of each iteration stage. Figure 15 shows the image mosaic formed by inserting the warped image into the reference image space.

7 Summary

The important result from this experimentation is that it is possible to register oblique image pairs using an iterative point correspondence method that has previously only been shown for 2-D rigid transformations in the image plane. The feature based approach relies on regions that can be segmented and common shape features between image views. It was shown that application of a global projective transformation is suitable if the surface being imaged can be approximated by a plane. An objective error criteria can be formed because we are using feature sets as opposed to raw image data. Furthermore, the reduction of the data from an image space to a feature set enables fast computation. The techniques applied here show promise for registering electro-optical images even if the images contain information disparity or vastly differing lighting conditions.

Frame 1
(Original Image)

Frame 2
(First warping based on
original feature point
correspondence)

Frame 3
(Second warping based on
point correspondence from
frame 2)

Figure 14: Frame 1 shows the original image to be aligned to the reference image. Frames 2 and 3 show the results of the projectively warped images after iteration 1 and iteration 2 of the algorithm. Note that after just two iterations the distorted image is fairly well aligned with the reference image.

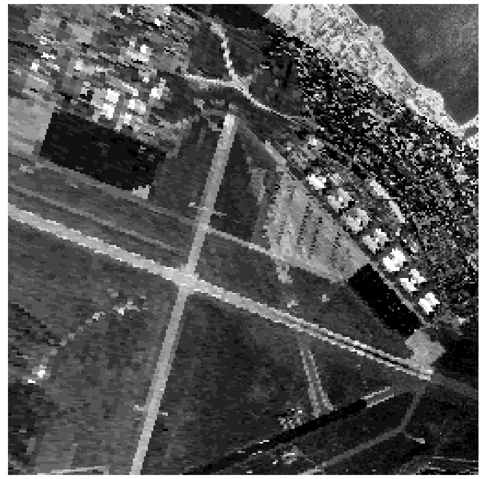

Figure 15: Original reference image versus generated mosaic image.

References

[1] Brown, Lisa Gottesfeld., "A Survey of Image Registration Techniques," ACM Computing Surveys, Vol. 24, No. 4, pp.325-376, December 1992.

[2] Growe, S. and Tonjes, R., "A knowledge based approach to automatic image registration," Proceedings International Conference on Image Processing, pp. 228-231 vol. 3, October 1997.

[3] Hsieh, Jun-Wei, et.al., "Image Registration Using a New Edge-Based Approach," Computer Vision and Image Understanding, Vol. 67 No. 2, pp. 112-130, 1997.

[4] Jain, Anil K., *Fundamentals of Digital Image Processing*, Englewood Cliffs: Prentice-Hall, 1989.

[5] Mann, S. and R.W. Picard, "Video Orbits of the Projective Group: A New Perspective on Image Mosaicing," M.I.T. Media Laboratory Perceptual Computing Section Technical Report No. 338.

[6] Ritter, Gerhard and Joseph Wilson, *Handbook of Computer Vision Algorithms in Image Algebra*, CRC Press. Boca Raton, New York, London, Tokyo, 1996.

[7] Wong, Robert Y., "Sensor Transformations" IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-7, No. 12, December 1977.

[8] Zhang,X., et.al.,"Automatic image-to-site model registration," IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol LVII+3588, 2164-7 vol.4, May 1996.

[9] Zheng,Qinfen and Chellappa, Rama,"Automatic registration of oblique aerial images," Proceedings ICIP, pp.218-222, Vol. 1, November 1994.